

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени
первого Президента России Б. Н. Ельцина»

ЗАДАНИЕ НА ПРОЕКТ

Екатеринбург
2024

Содержание

Перечень заданий по вариантам	3
-------------------------------------	---

Задание	4
---------------	---

Перечень заданий по вариантам

Студенты выбирают задание для работы в соответствии со своим **вариантом**, который определяется по последним двум **цифрам** студенческого билета. В случае если последние две цифры превышают **25**, то следует использовать номер варианта в виде целого остатка от деления последних цифр студенческого билета по модулю **25**.

К заданию прилагается архив с **25 txt**-файлами, внутри которых содержатся в один столбец различные значения реальных данных совершенно разных объемов, с абсолютно разными характеристиками. Студенты могут либо напрямую копировать значения из txt-файлов в свою рабочую «тетрадь», либо загружать значения из файлов с помощью различных функций (самый правильный вариант). В целях упрощения временная сетка исходных данных не предоставляется, и студенты могут создать ее сами (если того требуют вычислительные методы).

Все работы выполняются в соответствии с инструментами и методами, использованными в ходе выполнения практических заданий.

Задание

Результатом выполнения индивидуального задания является оформленный **отчет** в виде *Jupyter*-тетради, в котором должны быть представлены и отражены все нижеперечисленные пункты для **собственного варианта** временного ряда. Также необходимо иметь ввиду, что указанные ниже пункты не расположены в порядке конкретной последовательности методов для отчета, в которой их нужно применять, а просто перечислены в порядке их появления в лекциях/на занятиях. То есть применять пункты ниже следует творчески, используя различные функции там, где они могут пригодиться, а не просто копируя код из практических заданий. Кроме того, у методов могут быть выбраны **любые параметры**, но все они обязательно требуют пояснения своего выбора.

По итогу, с любыми данными нужно сделать следующее: разбить временной ряд на аддитивные компоненты (тренд + сезонные/циклические компоненты + шум) и построить прогноз данных своего варианта. При этом, для возможности проверки точности прогноза, следует делать изначально ретроспективный прогноз на известных значениях.

- 1) Проверьте временной ряд на свойство «стационарности». При необходимости – также используйте методы проверки статистических гипотез применительно к выделенным компонентам ряда, для пояснения своих решений по выбору параметров того или иного метода.
- 2) Используйте методы построения периодограмм и спектрограмм для подтверждения числа компонент, на которые Вы в дальнейшем будете разбивать исходный временной ряд. Также используйте спектрограммы для подтверждения «периодичности» выделенных компонент методами вейвлет-разложения или SSA.

- 3) При наличии частотной модуляции (или зависимости периода компонент от времени) в выделенных компонентах ряда, для подтверждения этого можно использовать метод мгновенной частоты на основе Преобразования Гильберта.
- 4) Для выделения **тренда** – используйте методы регрессии и/или методы сглаживания. Если вид исходных данных это позволяет – то сделайте и прогноз этих данных с помощью методов регрессии.
- 5) Разбейте исходные данные на аддитивные компоненты (тренд + сезонные/циклические компоненты + шум) с помощью **метода вейвлет-декомпозиции**. Базисный вейвлет, уровни декомпозиции, параметры '*mode*' – **любые**, но должны быть пояснения.
- 6) Разбейте исходные данные на аддитивные компоненты (тренд + сезонные/циклические компоненты + шум) с помощью **метода SSA**. Длина окна и способ группировки компонент – **любые**, но должны быть пояснения.
- 7) Постройте **прогноз данных** с помощью модели **SARIMAX, ARIMA** или других подобных, которые подойдут наилучшим образом для Вашего варианта данных. Не забывайте, что в некоторых случаях из исходных данных потребуется сначала устранить тренд и/или шум для того, чтобы такую модель в принципе удалось построить. Иногда можно построить прогноз этими моделями только для отдельных компонент, но не всего временного ряда в целом – об этом тоже следует задуматься. Для подтверждения выбора «сезонности» в моделях не забывайте про периодограммы. Оптимальность выбора параметров моделей SARIMAX и подобных можно доказывать через информационные критерии.

- 8) Постройте **прогноз данных** с помощью рекурсивного метода Сингулярного Спектрального Анализа **SSA-R**. При этом допускается сначала «очистить» исходные данные от шума и/или тренда с помощью других методов и только затем прогнозировать ряд.
- 9) Постройте **прогноз данных** с помощью нейронной сети любой конфигурации слоев, числа нейронов, типа слоев, типа сети (LSTM, GRU и т.д.). При этом допускается сначала «очистить» исходные данные от шума и/или тренда с помощью других методов и только затем прогнозировать ряд.

В любом случае – не забывайте, что задание творческое, а **реальные** данные у всех – абсолютно разные. Для одних лучше подойдут одни методы и параметры, а для других вариантов – другие. Это означает, что вполне будут зачтены даже те случаи, где прогноз не удался, или компоненты не построились, или ARIMA-модель вообще не получается. Но любой такой факт следует четко аргументировать, если аргументация достаточно убедительна – задание все равно будет зачтено.

Успехов Вам в выполнении этого задания!